Course Overview

- This course is designed to teach you how to design and engineer complete web-based business applications quickly.
- The course relies heavily on Java and Object-Oriented design techniques as practiced in the Application Engineering and Development class.
- This course builds on your previous experience by showing you how to use advanced server-side technologies to build significant web-based applications.
- Issues of page navigation, object persistence, high performance, multi-user access are addressed.
- Although students are expected to learn html and javascript skills these topics will not be covered in class.
- The student is expected to learn these skills on their own.

What you will learn

- Develop dynamic data-driven web applications using the Spring MVC, Hibernate ORM, and DAO.

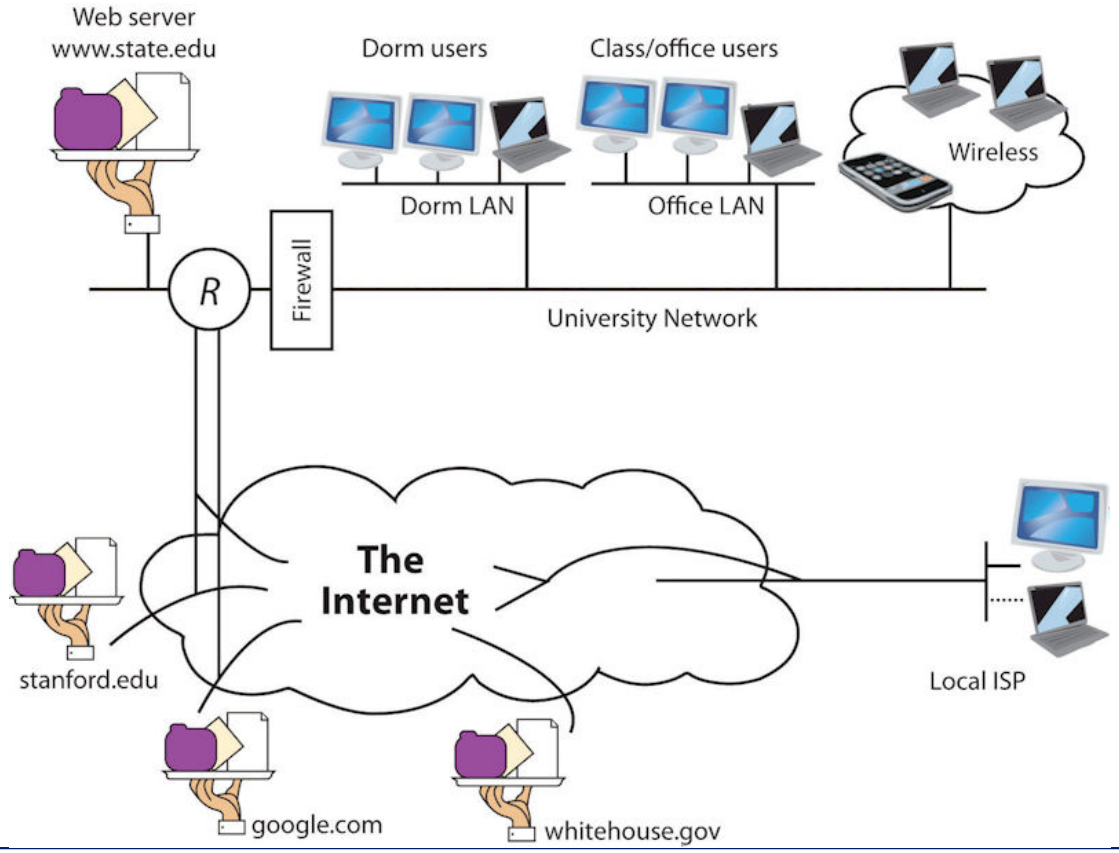
- Prerequisites:
  - ISYG100 (INFO5100) or permission of the Program Director.

What is computer networking?

- Networking is the practice of linking two or more computing devices together.

- What is the need for networking?

- The Internet is a worldwide, publicly accessible series of interconnected computer networks that transmit data by packet switching using the standard Internet Protocol (IP).
- It is a "network of networks" that consists of millions of smaller domestic, academic, business, and government networks, which together carry various information and services, such as electronic mail, online chat, file transfer, and the interlinked Web pages and other documents of the World Wide Web.



What is a Network Protocol?

- Network protocols defines a language of rules and conventions for communication between network devices.

HTTP

- provides a standard for Web browsers and servers to communicate. The definition of HTTP is a technical specification of a network protocol that software must implement.

- is a communications protocol used to transfer or convey information on the World Wide Web. Its original purpose was to provide a way to publish and retrieve HTML hypertext pages.

TCP/IP

- Transmission Control Protocol/Internet Protocol
- The most important protocol that makes Internet universality possible.
- It permits any computer at all to communuicate with the internet.
- This is rather like everyone in the world speaking one language.

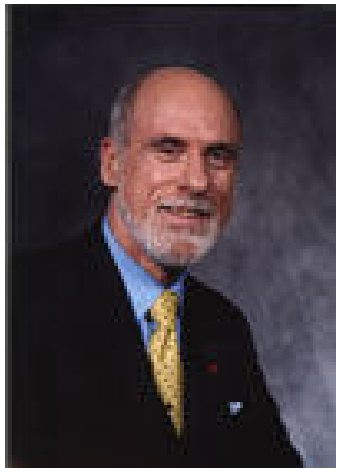
IP Address

- The entire purpose of IP is to provide unique global computer addressing to ensure that two computers communicating over the Internet can uniquely identify one another.
- An IP address is the logical address of a network adapter.
- The IP address uniquely identifies computers on a network.
- An IP address can be private, for use on a LAN, or public, for use on the Internet or other WAN.
- IP addresses can be determined statically (assigned to a computer by a system administrator) or dynamically (assigned by another device on the network on demand).

Who invented the Internet Protocol?

- In a research paper published in 1974, Vinton Cerf and Robert Kahn proposed a protocol they called "TCP".

- Cerf and Kahn didn't realize it at the time, but the protocol they invented would later become IP, the official network-layer protocol of the Internet.



IPv4

- IPv6 is the proposed successor to IPv4 whose most prominent change is the addressing.

- IPv4 uses 32-bit addresses
  - (~4 billion addresses)

- while IPv6 uses 128-bit addresses
  - (~3.4×10^{38} addresses).

- Although adoption of IPv6 has been slow, as of 2008, all United States government systems must support IPv6 (if only at the backbone level).

What is the size of today's standard (IPv4) IP address?

a) 4 bytes (32 bits)
b) 12 bytes (96 bits)
c) 15 bytes (120 bits)
d) 16 bytes (128 bits)

What is the size of the next-generation (IPv6) IP address?

- a) 4 bytes (32 bits)
- b) 12 bytes (96 bits)
- c) 15 bytes (120 bits)
- d) 16 bytes (128 bits)

IPv4 Vs IPv6

- **An IPv4 address contains 4 bytes (32 bits).**
  - 4,294,967,296 IPv4 addresses
- IPv4 addresses are written in dotted-decimal notation where a period (.) separates each byte
  - 10.0.0.255

- **An IPv6 address contains 16 bytes (128 bits)**
  - ~3.4×10^{38} IPv6 addresses
- IPv6 addresses are written in a hexadecimal format.
- 8 groups of 2-byte segments of the address are separated by colons (:)
  - E3D7:0000:0000:0000:51F4:9BC8:C0A8:6420

What does the following IP address belong to?

- http://129.10.32.100
- http://74.125.226.9
- http://54.239.17.6
- http://157.166.226.26

- Is there an easy way to remember IP addresses?

The Most Basic Use Of DNS Is To Translate Hostnames To IP Addresses

- It is in very simple terms like a phone book. For example, if you want to know the internet address of www.cnn.com, the Domain Name System can be used to tell you it is 157.166.226.26.
- Pre-eminently, DNS makes it possible to assign Internet destinations to the human organization or concern they represent, independently of the physical routing hierarchy represented by the numerical IP address.
- Because of this, hyperlinks and Internet contact information can remain the same, whatever the current IP routing arrangements may be, and can take a human-readable form (such as “www.coe.neu.edu") which is rather easier to remember than an IP address (such as 129.10.32.100).
- People take advantage of this when they recite meaningful URLs and e-mail addresses without caring how the machine will actually locate them.
- The Domain Name System (DNS) distributes the responsibility for assigning domain names and mapping them to IP networks by allowing an authoritative server for each domain to keep track of its own changes, avoiding the need for a central registrar to be continually consulted and updated.

Managing Domain Names

- InterNIC or Internet Network Information Center was the Internet governing body primarily responsible for domain name and IP address allocations until September 18, 1998 when this role was assumed by the ICANN body. It was accessed through the website internic.net
- ICANN (IPA /aɪkæn/) is the Internet Corporation for Assigned Names and Numbers. Headquartered in Marina Del Rey, California, ICANN is a California non-profit corporation that was created on September 18, 1998 in order to oversee a number of Internet-related tasks previously performed directly on behalf of the U.S. Government by other organizations, notably IANA.
- The tasks of ICANN include managing the assignment of domain names and IP addresses. To date, much of its work has concerned the introduction of new generic top-level domains. The technical work of ICANN is referred to as the IANA function; the rest of ICANN is mostly concerned with defining policy.

Purchasing Domain Names

- Thru Accredited Registrars

HTTP Server

- A computer program that is responsible for accepting HTTP requests from clients and serving them HTTP responses along with optional data contents, which usually are web pages such as HTML documents and linked objects (images, etc.).

UDP (User Datagram Protocol)

- UDP is a communications protocol that offers a limited amount of service when messages are exchanged between computers in a network that uses the Internet Protocol (IP).
- UDP is an alternative to the Transmission Control Protocol (TCP) and, together with IP, is sometimes referred to as UDP/IP.
- Like the Transmission Control Protocol, UDP uses the Internet Protocol to actually get a data unit (called a datagram) from one computer to another.
- Unlike TCP, however, UDP does not provide the service of dividing a message into packets (datagrams) and reassembling it at the other end.
- Specifically, UDP doesn't provide sequencing of the packets that the data arrives in. This means that the application program that uses UDP must be able to make sure that the entire message has arrived and is in the right order.
- Network applications that want to save processing time because they have very small data units to exchange (and therefore very little message reassembling to do) may prefer UDP to TCP.
- UDP provides two services not provided by the IP layer. It provides port numbers to help distinguish different user requests and, optionally, a checksum capability to verify that the data arrived intact.

TCP vs UDP

- Using TCP, the computer sending the data connects directly to the computer it is sending the data it to, and stay connected for the duration of the transfer.
- With this method, the two computers can guarantee that the data has arrived safely and correctly, and then they disconnect the connection.
- This method of transferring data tends to be quicker and more reliable, but puts a higher load on the computer as it has to monitor the connection and the data going across it.
- A real life comparison to this method would be to pick up the phone and call a friend. You have a conversation and when it is over, you both hang up, releasing the connection.

TCP vs UDP Cont'd

- Using UDP, the computer sending the data packages the information into a nice little package and releases it into the network with the hopes that it will get to the right place.
- What this means is that UDP does not connect directly to the receiving computer like TCP does, but rather sends the data out and relies on the devices in between the sending computer and the receiving computer to get the data where it should go properly
- This method of transmission does not provide any guarantee that the data you send will ever reach its destination.
- On the other hand, this method of transmission has a very low overhead and is therefore very popular to use for services that are not that important to work on the first try.
- A comparison you can use for this method is the plain old US Postal Service.
- You place your mail in the mailbox and hope the Postal Service will get it to the proper location.
- Most of the time they do, but sometimes it gets lost along the way.

TCP and UDP Ports

- As you know every computer or device on the Internet must have a unique number assigned to it called the IP address. This IP address is used to recognize your particular computer out of the millions of other computers connected to the Internet.
- When information is sent over the Internet to your computer how does your computer accept that information? It accepts that information by using TCP or UDP ports.
- An easy way to understand ports is to imagine your IP address is a cable box and the ports are the different channels on that cable box.
- The cable company knows how to send cable to your cable box based upon a unique serial number associated with that box (IP Address), and then you receive the individual shows on different channels (Ports). Ports work the same way.
- You have an IP address, and then many ports on that IP address. When I say many, I mean many. You can have a total of 65,535 TCP Ports and another 65,535 UDP ports.
- When a program on your computer sends or receives data over the Internet it sends that data to an IP address and a specific port on the remote computer, and receives the data on a usually random port on its own computer.
- If it uses the TCP protocol to send and receive the data then it will connect and bind itself to a TCP port. If it uses the UDP protocol to send and receive data, it will use a UDP port.
- Note that once an application binds itself to a particular port, that port can not be used by any other application. It is first come, first served.

Ports Cont'd

- In order for a web server to accept connections from remote computers, it must bind the web server application to a local port.
- It will then use this port to listen for and accept connections from remote computers.
- Web servers typically bind to the TCP port 80, which is what the http protocol uses by default, and then will wait and listen for connections from remote devices.
- Once a device is connected, it will send the requested web pages to the remote device, and when done disconnect the connection.
- On the other hand, if you are the remote user connecting to a web server it would work in reverse.
- Your web browser would pick a random TCP port from a certain range of port numbers, and attempt to connect to port 80 on the IP address of the web server.
- When the connection is established, the web browser will send the request for a particular web page and receive it from the web server.
- Then both computers will disconnect the connection.

Ports Cont'd

- Now, what if you wanted to run an FTP server, which is a server that allows you to transfer and receive files from remote computers, on the same web server.
- FTP servers use TCP ports 20 and 21 to send and receive information, so you won't have any conflicts with the web server running on TCP port 80.
- Therefore, the FTP server application when it starts will bind itself to TCP ports 20 and 21, and wait for connections in order to send and receive data.
- Most major applications have a specific port that they listen on and they register this information with an organization called IANA. You can see a list of applications and the ports they use at the IANA Registry.
- With developers registering the ports their applications use with IANA, the chances of two programs attempting to use the same port, and therefore causing a conflict, will be diminished.

Static vs. Dynamic Web Pages

- A static Web page is a Web page that always comprises the same information in response to all download requests from all users.

- When a URL representing a static page is received by a Web server, the server always responds to that request with the same set of HTML and associated Web content, regardless of the user's identity or the retrieval context.

Dynamic Web Pages

- Classical hypertext navigation occurs among "static" documents, and, for web users, this experience is reproduced using static web pages.
- However, web navigation can also provide an interactive experience that is termed "dynamic". Content (text, images, form fields, etc.) on a web page can change, in response to different contexts or conditions.
- There are two ways to create this kind of interactivity:
  - Using client-side scripting to change interface behaviors within a specific web page, in response to mouse or keyboard actions or at specified timing events. In this case the dynamic behavior occurs within the presentation.
  - Using server-side scripting to change the supplied page source between pages, adjusting the sequence or reload of the web pages or web content supplied to the browser. Server responses may be determined by such conditions as data in a posted HTML form, parameters in the URL, the type of browser being used, the passage of time, or a database or server state.
- The result of either technique is described as a dynamic web page, and both may be used simultaneously.

Dynamic Web Pages

- To adhere to the first definition, web pages must use presentation technology called, in a broader sense, rich interfaced pages.
- Client-side scripting languages like JavaScript or ActionScript, used for Dynamic HTML (DHTML) and Flash technologies, are frequently used to orchestrate media types (sound, animations, changing text, etc.) of the presentation.
- The scripting also allows use of remote scripting, a technique by which the DHTML page requests additional information from a server, using a hidden Frame, XMLHttpRequests, or a Web service.

- Web pages that adhere to the second definition are often created with the help of server-side languages such as PHP, Perl, ASP or ASP.NET, JSP, and other languages.

Amazon.com in 1995

Discussion

- Which site is more complex?
  - http://www.northeastern.edu
  - http://www.google.com

Dynamic Data-Driven Web Site

- Why do we need to execute code on the server?

Surfing the Internet

- When you surf the Internet, you basically request for a certain file located in a particular computer in the location you specify in the Uniform Resource Locator (URL).
- The computer where the file is stored is called the web server.
- This computer's main function is to serve anybody on the Internet who requests files it hosts.
- Because you never know when a user will visit and use your web application, your web server must be up and running all the time.

- When you click or type in a URL in the Location or Address box of your browser, the following things happen:
  - The client browser establishes a TCP/IP connection with the server.
  - The browser sends a request to the server.
  - The server sends a response to the client.
  - The server closes the connection.

HTTP Revisited

- HTTP is the protocol that allows web servers and browsers to exchange data over the web.
- It is a request and response protocol.
- The client requests a file and the server responds to the request.
- In HTTP, it's always the client who initiates a transaction by establishing a connection and sending an HTTP request.
- The server is in no position to contact a client or make a callback connection to the client.
- Either the client or the server can prematurely terminate a connection.
- For example, when using a web browser you can click the Stop button on your browser to stop the download process of a file, effectively closing the HTTP connection with the web server.

HTTP Requests

- An HTTP transaction begins with a request from the client browser and ends with a response from the server.

- An HTTP request consists of three components:
  - Method – URI -- Protocol/Version
  - Request headers
  - Entity body

HTTP request example

```
GET /servlet/default.jsp HTTP/1.1
Accept: text/plain; text/html
Accept-Language: en-gb
Connection: Keep-Alive
Host: localhost
Referer: http://localhost/SendDetails.htm
User-Agent: Mozilla/4.0
Content-Length: 33
Content-Type: application/x-www-form-urlencoded
Accept-Encoding: gzip, deflate
LastName=Ozbek&FirstName=Yusuf
```

HTTP Responses

- Similar to requests, an HTTP response also consists of three parts:
  - Protocol -- Status code -- Description
  - Response headers
  - Entity body

HTTP response example

```
HTTP/1.1 200 OK
Server: Microsoft-IIS/4.0
Date: Tue, 8 Jan 2008 13:13:33 GMT
Content-Type: text/html
Last-Modified: Mon, 7 Jan 2008 13:23:42 GMT
Content-Length: 112

<HTML>
<HEAD>
   <TITLE>HTTP ResponseExample</TITLE>
</HEAD>
<BODY>INFO625050 Web Tools</BODY>
</HTML>
```

System Architecture

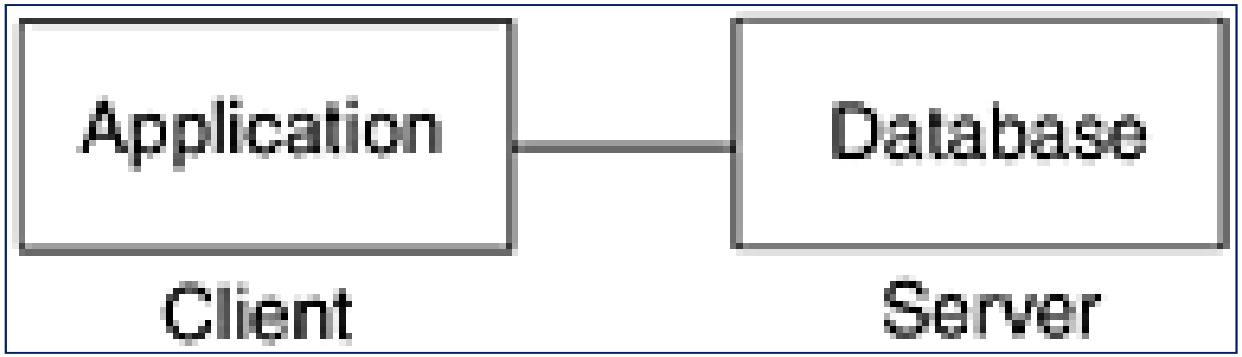
- A well-designed software application is partitioned into separate logical parts called layers.
- Each layer has a different responsibility in the overall architecture.
- These layers are purely abstractions, and do not correspond to physical distribution.
- **Typical layers in a software system are as follows:**
  - Presentation layer.
    - In this layer are parts that handle the user interface and user interaction.
  - Business logic layer.
    - This layer contains components that handle the programming logic of the application.
  - Data layer.
    - This layer is used by the business logic layer to persist state permanently.
    - This layer normally consists of one or more DBs where data is stored.

The Two-Tier Architecture

- A two-tiered application is a simple client-server application in which the processing workload falls onto the client computer's shoulders and the server simply acts as a traffic controller between the client and the data.

- The term "fat client" for this type of architecture is due to the bulk of processing requirements at the client side.

- In this architecture, the presentation layer and the business logic layer are hosted in one tier and the data layer is on another tier.



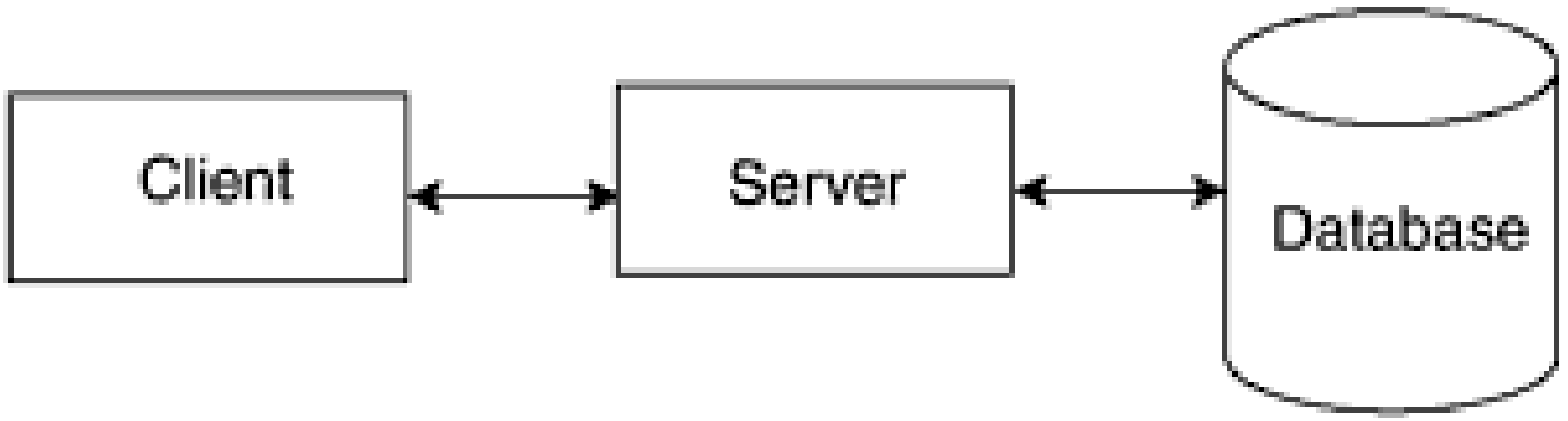
Problems with Two-Tier Architecture

- The drawback of this type of architecture is that it starts to pose problems as the number of clients increases.
  - Increased Network Traffic
    - Because each client has to make multiple requests for data from the server as all processing happens at the client side.
  - Increased Cost
    - Because each client needs a machine with sufficient processing power.
    - As the number of clients increase, the cost for providing client machines alone could be astronomical.
  - Maintenance Problem
    - Even a tiny change to the processing logic might require a complete rollout to the entire organization.
    - Even though the process can be automated, there are always problems with larger organizations because some users may not be ready for an upgrade, whereas others insist it be performed immediately.

The Three-Tier Architecture

- To overcome the problems in many client two-tiered applications, an application is broken up into three separate tiers, instead of two.

- The first tier contains the presentation layer, the second tier, or the middle tier, consists of the business logic layer, and the third tier contains the data layer.



Java Enterprise Edition

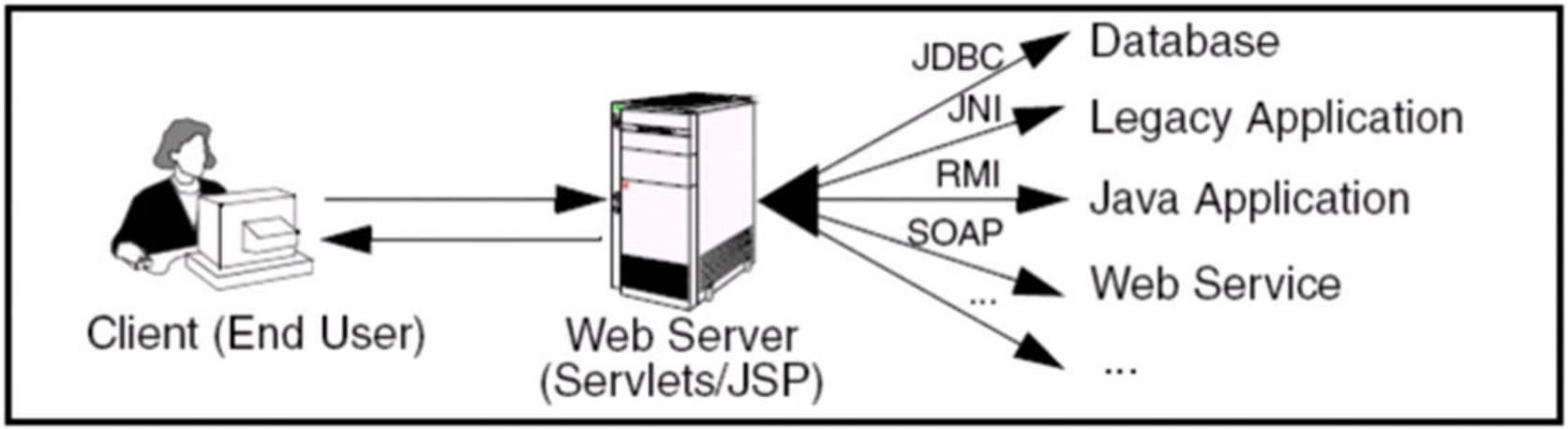
- First of all, J2EE is not a product.
- Rather, it is a specification that defines the contract between applications and the container.
- The container here refers to a standardized runtime environment, which provides specific services for components deployed in it.

- Java EE includes the following APIs
  - JSP
  - Servlets
  - EJB
  - JDBC
  - JTA (Java Transaction API)
  - JNDI (Java Naming and Directory Interface)
  - JMS (Java Message Service)
  - RMI (Remote Method Invocation)
  - JAX (Java API for XML)
  - SOAP and Web Services

Servlets/JSP

- Servlet and JSP technology has become the technology of choice for developing online stores, interactive Web applications, and other dynamic Web sites.

- Servlets are Java programs that run on Web or application servers, acting as a middle layer between requests coming from Web browsers or other HTTP clients and databases or applications on the HTTP server.



Discussion

- Why wait until the client requests the page and then have a program build the result?
- Why not just build the Web page ahead of time?”
- Yes, many client requests can be satisfied by prebuilt documents, and the server would handle these requests without invoking servlets.
- In many cases a static result is not sufficient, and a page needs to be generated for each request.
- There are a number of reasons why Web pages need to be built on-the-fly.

Why web pages need to be built on-the-fly?

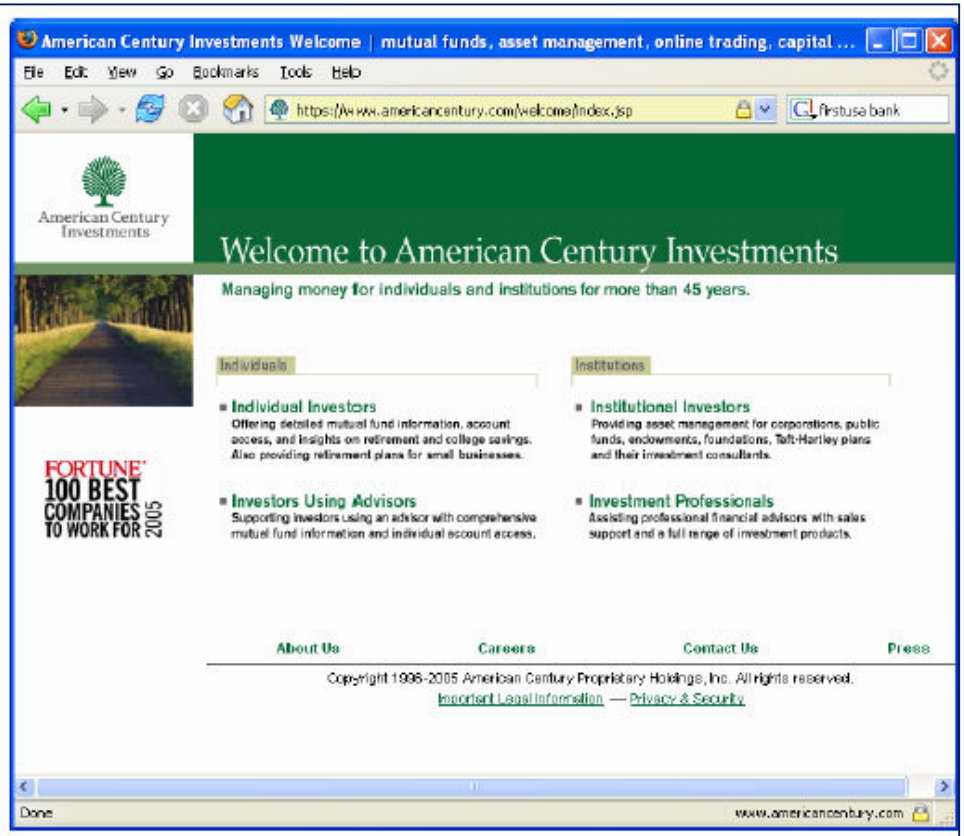
- **The Web page is based on data sent by the client.**
  - For instance, the results page from search engines and order confirmation pages at online stores are specific to particular user requests.
  - You don't know what to display until you read the data that the user submits.
- **The Web page is derived from data that changes frequently.**
  - If the page changes for every request, then you certainly need to build the response at request time.
  - If it changes only periodically, however, you could do it two ways:
    - you could periodically build a new Web page on the server (independently of client requests)
    - you could wait and only build the page when the user requests it.
  - The right approach depends on the situation, but sometimes it is more convenient to do the latter: wait for the user request. For example, a weather report or news headlines site might build the pages dynamically, perhaps returning a previously built page if that page is still up to date.
- **The Web page uses information from databases or other server-side sources.**
  - If the information is in a database, you need server-side processing.

JSP/Servlets in the Real World: Airlines

- Travelocity.com
- Orbitz.com
- HotWire.com
- Hotels.com
- CheapTickets.com
- National Car Rental
- Avis Car Rental
- Enterprise Car Rental
- Hertz Car Rental

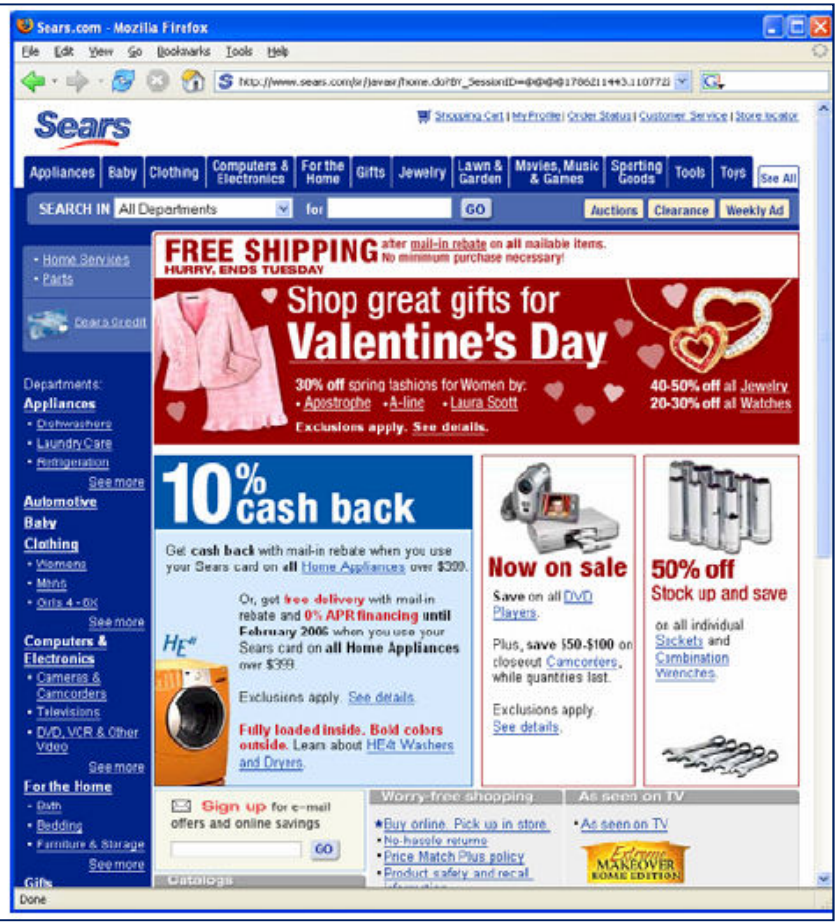
JSP/Servlets in the Real World: Financial Services

- American Century
- Vanguard
- Fidelity
- NY Stock Exchange
- First USA Bank
- Royal Bank of Scotland
- Banco Popular de Puerto Rico
- Bank of America
- China Construction Bank



JSP/Servlets in the Real World: Retail

- Sears.com
- Walmart.com
- SamsClub.com
- Macys.com
- llbean.com
- Kohls.com
- Ikea.com
- REI.com
- Longaberger.com
- Nike.com
- Polo.com

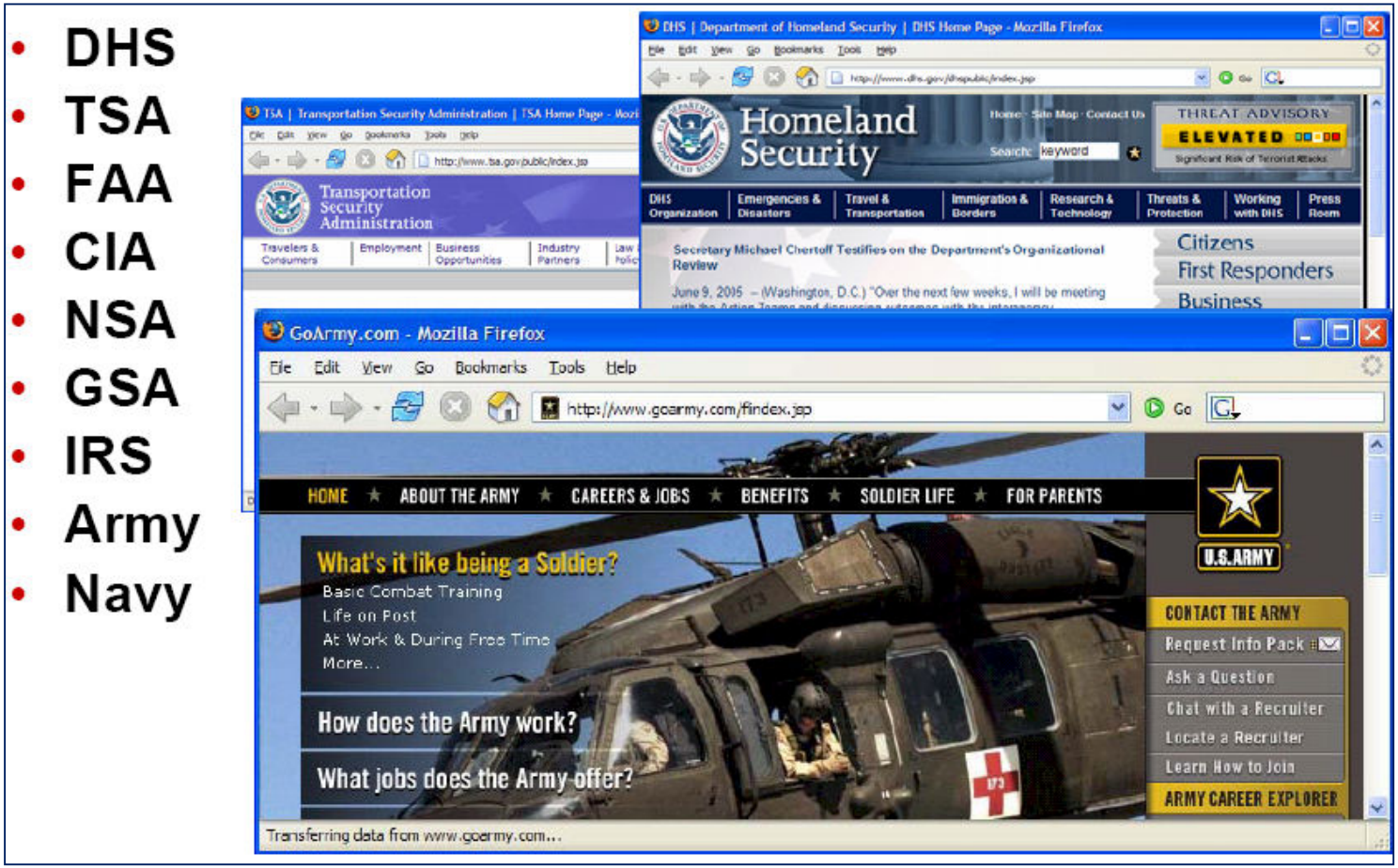


JSP/Servlets in the Real World: Entertainment

- WarnerBrothers.com
- Billboard.com
- E! (eonline.com)
- PBS.org

JSP/Servlets in the Real World: Military and Government

- DHS
- TSA
- FAA
- CIA
- NSA
- GSA
- IRS
- Army
- Navy



JSP/Servlets in the Real World: Sports

- Baltimore Orioles
- Baltimore Ravens
- Washington Redskins
- Washington Nationals
- Major League Baseball
- NHL.com
- Nascar.com

JSP/Servlets in the Real World: Search/Portals

- Most of Google
- All of Ebay
- netscape.com
- excite.com
- dice.com
- hi5
- Paypal